

## DOCUMENT RESUME

ED 326 196

IR 014 736

AUTHOR Kell, Diane; And Others  
TITLE Evaluating Technology in Schools: Implications of a Research Study.  
PUB DATE 90  
NOTE 23p.; Paper presented at the Annual Meeting of the American Educational Research Association (Boston, MA, April 16-20, 1990). For related papers, iss IR 014 735-738. Figure 1 will not reproduce well due to small, filled type.  
PUB TYPE Speeches/Conference Papers (150) --  
EDRS PRICE MF01/PC01 Plus Postage.  
DESCRIPTORS \*Adoption (Ideas); Change; \*Computer Assisted Instruction; Computer Software; Educational Technology; Elementary Education; \*Evaluation Methods; \*Evaluation Research; \*Instructional Innovation; Language Arts; Microcomputers; \*Program Implementation

## ABSTRACT

This report provides an overview of the issues related to a recently completed study of the use of computers in primary classrooms in six school districts and some special program classes, i.e., K-2 English-as-a-Second-Language and bilingual classes at one study site and Chapter 1 pullout classes for grades K-5 at another site. The sites used were demonstration sites for a new program of computer-based language arts instruction known as the Apple Software Series: Early Language (ALS-EL). Discussion of the development of the research context, purposes, and design includes consideration of issues related to: (1) assessing the implementation of educational innovations (an understanding of the process of change was emphasized and the change experience for individual teachers was examined); (2) assessing student performance in reading and writing (the Sulzby Writing and Rereading Task was adapted, and the Test of Early Reading Ability by Pro-Ed was chosen); and (3) assessing the effects of technology in classrooms (special attention was paid to the amount of funding allowed for the study, the amount of time necessary for the study, and the development of instruments to measure the implementation of an innovation). Strengths and weaknesses of the study are discussed, and implications for future research on computers in education are noted. It is suggested that a series of site-specific studies be pursued in which site-specific evaluation strategies be developed. It is concluded that operational variations and levels of implementation quality are important reminders of the need to allow time for major changes to take place before judgement about the innovations is passed. (34 references) (DB)

\*\*\*\*\*  
\* Reproductions supplied by EDRS are the best that can be made \*  
\* from the original document. \*  
\*\*\*\*\*

ED326196

U.S. DEPARTMENT OF EDUCATION  
Office of Educational Research and Improvement  
EDUCATIONAL RESOURCES INFORMATION  
CENTER (ERIC)

X This document has been reproduced as  
received from the person or organization  
originating it.

☐ Minor changes have been made to improve  
reproduction quality.

• Points of view or opinions stated in this docu-  
ment do not necessarily represent official  
OEI position or policy.

EVALUATING TECHNOLOGY IN SCHOOLS:  
IMPLICATIONS OF A RESEARCH STUDY

by

Diane Kell, Glen Harvey,  
and  
Nancy Gadzuk Drexler

April 1990

"PERMISSION TO REPRODUCE THIS  
MATERIAL HAS BEEN GRANTED BY

\_\_\_\_\_  
M. Simonson

\_\_\_\_\_  
TO THE EDUCATIONAL RESOURCES  
INFORMATION CENTER (ERIC)."

IR 014736

This document has been reproduced as  
received from the person or organization  
originating it.

Minor changes have been made to improve  
reproduction quality.

Points of view or opinions stated in this docu-  
ment do not necessarily represent official  
OERI position or policy.

## EVALUATING TECHNOLOGY IN SCHOOLS: IMPLICATIONS OF A RESEARCH STUDY

Diane Kell, Glen Harvey, and Nancy Gadzuk Drexler\*

It is not easy to gather evidence about the effects of computers on teaching and learning, especially when dealing with complex, naturally occurring variations in the real world of schools. So many variables affect classroom learning and instruction that it is quite nearly impossible to control for them when studying computer applications in real settings under non-laboratory conditions.

Recent reviews of the research on educational computing - e.g., Clark (1985); Becker (1987); Roblyer, Castine, and King (1988); and Krendl and Williams (1989) -- generally have concluded that the results of the studies completed thus far have been mixed or else must be discounted because of poor research design. Problems cited in these reviews include: lack of a control group for comparisons, non-random assignment of students to treatment vs. control, absence of comparable instructional opportunities for the control group, short-term studies lacking a longitudinal orientation, failure to address the issue of novelty effects, and failure to account for possible teacher quality differences between treatment and control groups.

Unfortunately, these reviews fail to raise the equally important issues of variation in *how* computers and computer-based instructional programs are used by individual teachers, whether the computer applications being studied are indeed *implemented* as intended by the developers, and the developmental process that individual teachers must go through to become knowledgeable and effective users of educational technology.

Moreover, these reviews focus almost entirely on the measurement of student performance outcomes, such as reading and math achievement. Some advocates for educational computing -- e.g., Weir (1989); Laboratory of Comparative Human Cognition (1989) -- are just as interested in the role computers might play in transforming the teaching and learning process. Others who fully support their vision -- e.g., Cohen (1988); Cuban (1989); LaFrenz & Friedman (1989); Pearlman (1989) -- doubt whether transformations in teaching and learning can occur without accompanying social, cultural, political, and organizational changes in school environments. Whether you believe computers can change schools, or whether you believe they can support changes that are already envisioned and supported in the larger school context, it still seems important to assess these changes as potential outcomes of computers in classrooms.

---

\*The research reported in this paper, presented at the 1990 annual conference of the American Educational Research Association, was supported by Apple Computer, Inc., and conducted by The Regional Laboratory for Educational Improvement of the Northeast and Islands, 300 Brickstone Square, Suite 900, Andover, MA, 01810. We also want to acknowledge the valuable contributions of our colleagues: Lynne Krensky, Robert Thomas, Felisa Tibbitts, Kirsten Schaetzel, Carol Red, and Elizabeth Wertz. However, the views presented in this paper are our own and are not necessarily those of our colleagues, The Regional Laboratory, or Apple Computer, Inc.

While not actually viewing computers as tools for transforming schools, other advocates for educational computing -- e.g., Compaine (1983); Starr (1983) -- have emphasized the role of computers in developing new levels and standards of literacy among America's young people, to match the needs for literacy of all kinds in our increasingly technological society. Others who support this notion -- e.g., Maddux (1988) -- point out the great differences among types of computer applications, many of which simply transfer old methods of instruction, such as basic skills drill-and-practice, to a new medium.

After noting that research into nearly all computer applications has produced some positive and some negative results, Maddux declares:

*I believe this exposes a common myth about computers, and also says a great deal about the kinds of research that should be conducted in the future. The exposed myth is the common belief that all we need to do is place a computer and a child in the same room and wonderful things will happen. That is no more true for computers than it is for books. If it were true, all or nearly all findings would be positive. Obviously, there are other, more powerful variables than mere use of the computer. How the computer is used must constitute an important set of variables. Therefore, I believe we need no more research intended to test computer vs. non-computer teaching. Future research should instead be aimed at finding the most effective ways of using computers at given grade levels, with given kinds of students, or in given subjects (p.8).*

One of the strengths of our study is that we started with a fairly realistic understanding of the complex nature of our task. That is not to say we avoided all the pitfalls in our attempts to assess the effects of educational computing in classrooms. Still, we think our design includes advances on some methodological fronts that should be considered in future research.

This paper, then, will provide an overview of our research context, purposes, and design. It also provides an opportunity to reflect on the strengths and weaknesses of our study and to draw implications from our experience for future research on computers in education.

### Study Background

The research described in this paper was funded by Apple Computer, Inc., and conducted by The Regional Laboratory for Educational Improvement of the Northeast and Islands during the 1988-89 school year. Our sites were demonstration sites for a new program of computer-based language arts instruction developed by Apple Computer and marketed for use in primary classrooms. Known as the Apple Learning Series: Early Language (ALS-EL), this program was designed with the intention of demonstrating the power and flexibility of the computer as a tool for language arts instruction inside the regular classroom, in the hands of the classroom teacher.

The sites for the demonstration were selected by Apple Computer in the spring of 1988. They included ten schools in six, very large, mostly urban school districts dispersed across the western,

midwestern, and southern regions of the country. Most of the 48 participating classes were regular kindergarten or first-grade classes. However, as shown in Table 1, the study also involved some special program classes, including kindergarten, first-, and second-grade ESL or bilingual classes in one study site, and Chapter 1 pullout classes for kindergarten through fifth-grade students in another study site. As discussed further below, 14 kindergarten, first-, and second-grade classrooms were involved as comparisons in three study sites.

Apple Computer gave each ALS-EL classroom the same configuration of hardware and peripheral devices to support the ALS-EL program, including three computers and a printer, two speech synthesizers, a touch screen, and a keyboard designed for children.<sup>1</sup> In addition, each teacher received the ALS-EL bundle of early language software. These software components, ranging from phonics and skills practice programs to word processing and graphics arts software for children, represent truly diverse approaches to early language development as well as quite different applications of the interactive potential of computer technology.<sup>2,3</sup>

Apple Computer packaged these programs together to give teachers the resources from which they could create a multifaceted, integrated language learning experience for children, consistent with their own language arts approaches and objectives. While different approaches to early language development might be represented in these programs, the ALS-EL developers expected that most teachers would find the full range of resources helpful in meeting the variety of student needs in their classrooms. To help guide them in creating their own applications using ALS-EL software, the developers prepared a teacher's manual. The manual describes the different software programs, shows the correspondence between the programs and typical K-2 language arts objectives, suggests ways of organizing students' work on computers, and provides sample lesson plans incorporating a variety of learning activities on and off the computers. In addition, Apple Computer provided one full day, or two half-days, of initial inservice training for the ALS-EL teachers in each site.<sup>4</sup>

In September 1988 Apple Computer funded The Regional Laboratory to study the effects of the ALS-EL program, and of classroom computing in general, on teachers and students, and on teaching and learning, in the six demonstration sites. During the first several months of the project, we developed a study design and *simultaneously* began efforts to collect baseline data on project classrooms. To capture the baseline data, we made our first field visits to study sites in October 1988. We visited them again in January, March, and May 1989.

Our study goals included assessing and portraying the effects of ALS-EL and of computers in general on teachers and students and on teaching and learning in the early elementary grades, and identifying innovative instruments, methods, and models for conducting future building-based research on the uses and impacts of technology in classrooms. Our research questions focused on describing and understanding the process of implementing ALS-EL in classrooms, the effects of

**TABLE 1**  
**SAMPLE OF THE ALS-EL EVALUATION PROJECT**

| Type of Class                       | Apple Classes        |                       |                              | Comparison Classes   |                       |                             |
|-------------------------------------|----------------------|-----------------------|------------------------------|----------------------|-----------------------|-----------------------------|
|                                     | Number of<br>Classes | Number of<br>Teachers | Number of<br>Focal* Students | Number of<br>Classes | Number of<br>Teachers | Number of<br>Focal Students |
| <b>Kindergarten:</b>                |                      |                       |                              |                      |                       |                             |
| Regular                             | 14                   | 9                     | 84                           | 5                    | 4                     | 30                          |
| ESL/Bilingual                       | 2                    | 2                     | 12                           | 1                    | 1                     | 6                           |
| Chapter 1 Pullout                   | 1                    | 1                     | 6                            | -                    | -                     | -                           |
| Chapter 1 All Day                   | 1                    | 1                     | 6                            | -                    | -                     | -                           |
| <b>First Grade:</b>                 |                      |                       |                              |                      |                       |                             |
| Regular                             | 16                   | 16                    | 96                           | 5                    | 5                     | 30                          |
| ESL/Bilingual                       | 2                    | 2                     | 12                           | 1                    | 1                     | 6                           |
| Chapter 1 Pullout                   | 1                    | 1                     | 6                            | -                    | -                     | -                           |
| High Intensive<br>Language Training | 1                    | 1                     | 6                            | 1                    | 1                     | 6                           |
| <b>Second Grade:</b>                |                      |                       |                              |                      |                       |                             |
| Regular                             | -                    | -                     | -                            | -                    | -                     | -                           |
| ESL/Bilingual                       | 2                    | 2                     | 12                           | 1                    | 1                     | 6                           |
| Chapter 1 Pullout                   | 2                    | 2**                   | 9                            | -                    | -                     | -                           |
| <b>Other:</b>                       |                      |                       |                              |                      |                       |                             |
| Early Childhood Program             | 1                    | 1                     | 6                            | -                    | -                     | -                           |
| Chapter 1 Pullouts                  |                      |                       |                              |                      |                       |                             |
| Third Grade                         | 2                    | 2                     | 12                           | -                    | -                     | -                           |
| Fourth Grade                        | 2                    | 2**                   | 9                            | -                    | -                     | -                           |
| Fifth Grade                         | 1                    | 1                     | 6                            | -                    | -                     | -                           |
| <b>TOTALS</b>                       | <b>48</b>            | <b>43</b>             | <b>282</b>                   | <b>14</b>            | <b>13</b>             | <b>84</b>                   |

\*Teachers were asked to select six "focal" students per class, who were tracked more intensively. Focal students were selected randomly through a stratified procedure resulting in an even distribution across high, medium, and low ability groups and between girls and boys.

\*\*Includes one part-time teacher; the sample therefore includes nine (rather than twelve) focal students.

the program on teacher attitudes and behaviors and on the classroom environment, and the effects of the program on student performance in the language arts, and on student attitudes and behavior.

As an explanatory model, we developed the conceptual framework shown in Figure 1. As seen in this figure, we hypothesized complex interrelationships among: (1) school climate and culture; (2) organizational supports and constraints for teaching practices involving the use of computers; (3) teachers' attitudes, beliefs, skills, and knowledge, especially relating to early language development, instructional computing, and general orientations to teaching and learning; (4) the teachers' customary approaches to classroom management, classroom design, and instruction; (5) the developmental stages and operational variations of ALS-EL implementation; (6) student background characteristics; and (7) student outcomes.

### Study Issues and Strategies

In developing our study design, we drew on several fields of educational inquiry, most importantly, the learnings from research on the implementation of educational innovations, assessment of student performance in the language arts, and assessment of instructional technology.

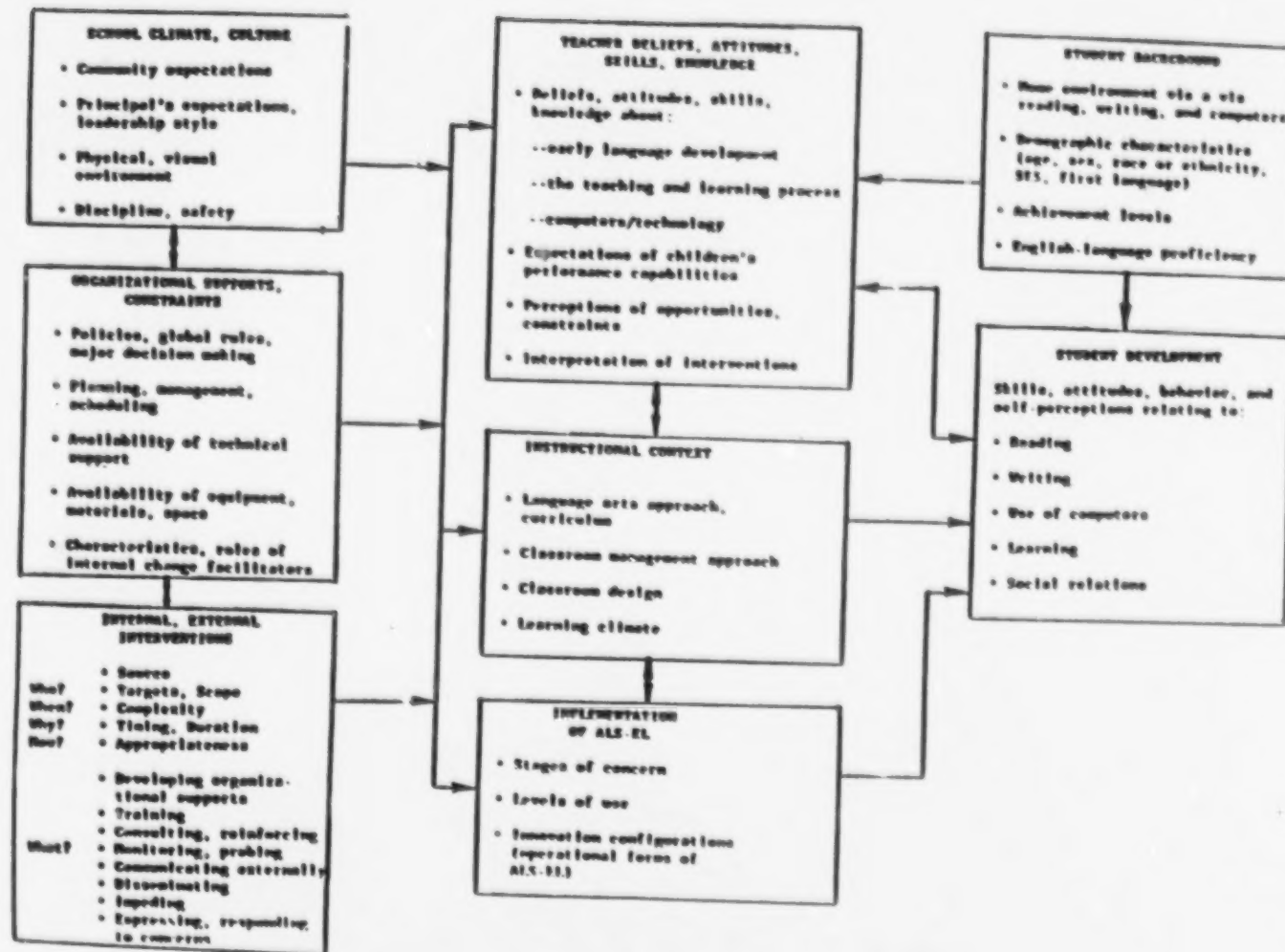
#### Issues Related to Assessing Implementation of Educational Innovations

From implementation research, we developed our emphasis on understanding the process of change from the teacher's perspective as well as understanding how teachers were actually using ALS-EL in their classrooms before attempting to determine what differences the program made in their practice or in student outcomes. These ideas are stressed in the Concerns-Based Adoption Model (CBAM) developed by Gene Hall and his associates at the Research and Development Center for Teacher Education in Austin, Texas. We understand that "*change is a process, not an event*" and that "*change takes time*" (Hall & Hord, 1987). Teachers have to work through a multitude of personal and management concerns before they can become truly comfortable and proficient in their use of computers in the classroom. Therefore, we did not expect to see major impacts on teaching and learning during the first year following the introduction of computer technology to classrooms. In fact, we realized it could be several years before changes would occur to affect student outcomes in a significant way.

We also expected to find many variations in how teachers chose to adapt the ALS-EL resources to fit their customary teaching approaches, curriculum, and methods of classroom management; and we felt it was important to understand and document these differences so that they could be related to student outcomes at whatever time they were assessed. Research beginning with the Rand Change Agent Studies (Berman & McLaughlin, 1975; Berman & McLaughlin, 1978) made it clear that we should expect to find "*mutual adaptations*" of the innovation and of the user's customary practice. However, we wondered to what extent computers would stimulate changes in customary teaching approaches, curriculum, and methods of classroom management, and to what extent the use of computers would be constrained by these same features of teachers and classrooms. For

Figure 1

CONCEPTUAL FRAMEWORK



example, what would happen if computer learning stations were brought into classrooms where instruction had always been whole-class and teacher-directed, where the children always sat in rows in desks, where learning was generally passive, and where the only interactions were between teacher and students? How would the teacher manage the children's use of computers? What kinds of computer applications would the teacher choose to use (or not use)? Would teachers begin experimenting with other small-group or independent, centers-based learning activities? And what if the teacher stressed a "phonics-first" approach to language arts instruction? Would she use the word processing software that comes with the package, if before this time she believed children could not write until they had learned to spell? Would the children's use of computers affect the teacher's views about how children learn or what they are capable of learning and doing?

To address these and other implementation issues, we made extensive use of CBAM concepts, as adapted for prior work by The Regional Laboratory, in our research design. Most importantly, we attempted to assess the several dimensions of the change process as it is experienced by individual teachers -- i.e., Stages of Concern, Levels of Use, and Innovation Configurations -- with varying degrees of fidelity to the CBAM diagnostic approach. We included the Stages of Concern questionnaire in our surveys of ALS-EL teachers in January and May 1989. Although we did not measure Levels of Use, we tried to capture some of this dimension in our own assessment of the "levels of implementation quality." We also developed an "innovation components checklist" that could be used to describe the operational variations of ALS-EL, or Innovation Configurations, actually in use in classrooms.<sup>5</sup>

Our field researchers spent a fairly large amount of time in schools collecting unstructured, or loosely structured, interview and observation data, in addition to completing these rather systematic applications of the CBAM concepts. We completed a total of 32 to 40 person-days of field work in each district site during the 1988-89 school year. During each of our four visits, we conducted open-ended, semi-structured interviews with ALS-EL teachers (about 45 minutes to an hour with each teacher) and with school administrators (generally one to two hours for an administrator) and observed approximately two hours of language arts instruction and 45 minutes of children's computer use in each classroom. The interviews and observations from each visit were documented in the researchers' field notes. Near the end of the year, the field researchers completed the "innovation components checklist" and an "implementation rating sheet" which was used in developing our measure of implementation quality in different classrooms. Having two researchers on each field team, with overlap between teams, helped to ensure some consistency in rating decisions across sites.

The changes in ALS-EL classrooms, perceived impacts on students and teachers, and factors affecting ALS-EL use were also discussed at periodic two-day meetings of the entire research staff. During these meetings, we developed, amended, and refined analytic propositions relating to each of our study questions. At the end of the year, following the advice of qualitative researchers such as Matthew Miles and Michael Huberman (1984), we tested these propositions by asking each field researcher to complete a chart of the evidence in their field reports, including negative evidence as well as evidence supporting the proposition for each teacher they had studied. For some

propositions, we also constructed contingency tables and other kinds of data display that helped us in drawing our conclusions.

### Issues Related to Assessing Student Performance in Reading and Writing

Even though we did not expect to find student impacts in the first year of ALS-EL use, our study purposes included identifying innovative instruments, methods, and models for conducting future building-based research on the uses and impacts of technology in classrooms.<sup>6</sup>

Traditional tests of student ability have come under attack recently, particularly as they are used at the elementary-level. Beyond the general kinds of problems attributed to traditional standardized tests (see, for example, Shepard, 1989), there is increasing concern about the inappropriate uses of standardized tests in assessments of very young children (Gold, 1988; Meisels, 1989). Furthermore, many observers of children's literacy development are noting the lack of correspondence between traditional standardized tests and new conceptions of reading and writing and children's development in these areas (see Valencia et al., 1989; Edelsky & Harmon, 1988; Teale, 1988).

Among the criticisms are these reported by Edelsky and Harmon:

*A study of 1,000 children [Altwerger & Resta, 1986] showed no particular relationship between their actual reading and their scores on the California Test of Basic Skills. Some children scored high but read poorly; others scored high and read well; some low scorers read well and others didn't . . . .*

*. . . . The reason correlations are high between test scores and school achievement is because both the tests and the classwork require the same thing -- simulations of reading and writing (p.159).*

For this reason, we experimented with a number of new approaches to assessing literacy development among children in primary classrooms. We used an adaptation of the Writing and Rereading Task developed by Elizabeth Sulzby and her associates at the University of Michigan to capture the emergent literacy skills of very young children. For a measure of reading ability, we chose the Test of Early Reading Ability published by Pro-Ed in Austin, Texas, because of its emphasis on finding meaning in print. We collected and analyzed writing samples that are the product of routine classroom activities, following procedures developed by Virginia Resta at the University of New Mexico.

The Sulzby Writing and Rereading Task was selected for the study on the advice of some who had been involved in developing the ALS-EL program. They believed that Sulzby's conceptualization of children's emergent literacy behaviors matched the instructional philosophy of the ALS-EL program quite well. For example, Sulzby questions the traditional hierarchical view of children's early writing development, which suggests that children give up earlier forms of writing as they adopt new ones that more closely approximate conventional writing. Sulzby argues (with Teale,

1985, p.10) that children "*simultaneously hold a number of hypotheses about writing in a repertoire. That is, the same child may use scribble for one story, letter strings for another, write his name conventionally, and write a list of words in a mix of conventional and invented spellings.*" She views reading and writing as inextricably linked, mutually supportive tasks, especially in children's early literacy development; and she has discovered that one way to understand children's development as readers and writers is to observe them composing, reading, and then rereading their own writing.<sup>7</sup> We used her method with selected "focal" students in ALS-EL classrooms during our October and May site visits to all project schools.<sup>8</sup> We also collected Sulzby samples in January for focal students in three district sites. In these three sites, Sulzby samples were collected for focal students in comparison classrooms as well, but only during the October and May visits.

The Test of Early Reading Ability (TERA) was recommended by several people whom we consulted in our search for new measures of literacy development in young children. Several features of the TERA were considered in our decision to use this instrument. They included an emphasis on finding meaning in print, which was very important to some teachers in our study; procedures for administering the TERA to individual students one at a time, which would promote a more relaxed and personalized environment for the students who would be tested; and the availability of norming tables that could be used to equate TERA scores with scores on other standardized tests. We trained teachers to administer the test during our October field visits. They then administered the test individually to approximately 12 children per class in October-November 1988 and April-May 1989. The tested children included the six focal students per class, representing high, medium, and low language arts ability, plus six additional, randomly selected students per class.

Virginia Resta helped us to develop procedures for collecting and analyzing samples of the writing that children produce in their classrooms as part of their normal learning experiences. Our procedures included asking teachers to send us samples for all students in their classrooms at the beginning and end of the year, and for focal students more often during the year. We began by asking for weekly samples from focal students but later decided to request no more than one sample per student for each month. The teachers also were asked to attach a label to each sample, identifying the student and indicating the context for the sample, such as whether the topic was assigned or self-selected. They were also asked to provide a transcript for any sample where the child's invented spelling might not be easily interpreted by coders at The Regional Laboratory.

Because our study did not continue beyond the first year of implementation, it is hard to evaluate the utility of these procedures for assessing the effects of technology on children's language or literacy development. We believe all three assessment procedures are valuable for assessing some types of outcomes. However, even within the area of language and literacy development, they do not cover the wide variety of outcomes that some people have suggested may result from computer use in classrooms.

## Issues Related to Assessing the Effects of Technology in Classrooms

We have already mentioned the frustration that many reviewers of research on computers express about the quality of the research evidence on the effects of computers in education. In his preface to Roblyer, Castine, and King's (1988) review, Maddux warns of a growing backlash of public and professional opinion against educational computing, which he believes may be *"fueled by . . . the unrealistically extravagant and unsubstantiated claims that have been made by some educational computing advocates."* Roblyer, Castine, and King themselves state that the major investment required to expand the use of computers in schools *"would be so great -- especially coming, as it does, at a time of declining resources for education -- that it requires some objective information to justify the expense and effort."*

We mentioned earlier two kinds of problems cited by reviewers of research on computers in schools. One set of problems has to do with the poorly controlled experimental conditions of most research studies of the effects of computers on student performance outcomes: lack of a control group for comparisons, non-random assignment of students to treatment vs. control, absence of comparable instructional opportunities for the control group, short-term studies lacking a longitudinal orientation, failure to address the issue of novelty effects, and failure to account for possible teacher quality differences between treatment and control groups.

For a number of reasons, we, too, did not perform a tightly controlled experimental study. The demonstration sites were selected, computers were placed in classrooms, and class lists were completed following the districts' routine student assignment procedures before we became involved in the project. Hence, no random assignment of computers to classrooms, or students to treatment, was possible. In all but one of the five sites where computers were placed in regular classrooms, they had been given to *every* kindergarten and first-grade classroom in the school, meaning that we would have had to remove the program from some classrooms if we were to set up experimental conditions within the participating schools.

When we began planning the study in September, we decided to recruit comparison classes in other schools in the same districts. With authorization from Apple Computer, Inc., we promised computers for these classrooms after completion of the study. However, because of our late start, we had very little time to select matched comparisons carefully and systematically. Instead, we relied on district administrators to select these sites based on criteria that we had developed, i.e., similar student demographics, similar teaching staffs, similar language arts programs. We ended up with comparison classes in three sites. In one of these sites -- the one where all classrooms were ESL or bilingual classrooms -- the comparison classes were at the same grade levels in the same school. In two other sites, the comparison classes were in a different school in the same general neighborhood of the same district.

We proposed a three-year study, but we knew from the start that funding for the second and third years was not assured. In fact, changing priorities within Apple Computer, Inc., left us without funding after one year. Thus, we were unable to continue the study to collect measurements over time. In our minds, this is the greatest weakness of our study. We collected huge amounts of

baseline data on student outcomes, but without subsequent years of data, their value is greatly diminished. In hindsight, it seems we should have devoted less of our study resources to this aspect of our study and thereby lessened the burden on teachers as well.<sup>9</sup>

A second set of problems with prior studies is that they have often neglected the important issues of variation in *how* computers are used, whether applications being studied are indeed *implemented* as intended by the developers, and whether teachers have progressed through the predictable stages of change to become effective users of educational technology before the applications were assessed. If we do not have an accurate picture of what has been implemented, we have very little chance of understanding what uses of technology are most effective. In addition, innovations may be judged to be ineffective, when in reality they are not fully implemented.

These are issues we considered carefully in our research. As mentioned above, we devoted a great deal of attention to describing specific operational variations of the ALS-EL program in use as well as assessing teachers' progress in becoming effective users of technology.

But, still, describing the many variations that could occur and relating these to student outcomes was a great challenge for us. As mentioned above, the ALS-EL program included a variety of resources that teachers were expected to use flexibly to create computer-based learning experiences that would fit their customary teaching approaches, curriculum, and methods of classroom management. Improved literacy outcomes were expected to result from teachers' judicious use of the ALS-EL resources to meet the individual needs of students in their classrooms.

While the program's flexibility may be seen as a strength, it complicates the task for evaluators. There was no program "archetype" against which we could assess the acceptability of the many variations in program use across classrooms. We needed a year of exploratory study just to identify the important variations that could occur. We would need to spend even more time in discussions with teachers, administrators, and program developers to develop a profile of ideal, acceptable, and unacceptable practices that could provide benchmarks for assessing the quality of program implementation in individual classrooms. Indeed, because of the wide variations in approach to language arts instruction, we should probably have to develop several different profiles representing different instructional approaches.

Finally, because of the breadth of curricular objectives addressed by the various software components, we could look for student impact in almost any area of the language arts curriculum and even beyond -- from children's ability to identify and form letters, to higher level reading comprehension, to creative thinking and problem solving, as just several examples.

Because different teachers choose to use different computer applications for different learning objectives, we could not possibly expect to find every type of impact in all classrooms. Even if all teachers were to become "high" implementers, we should still expect to find different outcomes in classrooms where teachers were emphasizing different objectives, especially since variations would occur not only in teachers' use of ALS-EL but throughout the language arts curriculum.

Very likely, the outcomes would remain hidden in analyses that aggregated them across classrooms where different objectives were emphasized.

In our own study, coming as it did in the first full year of computer use in classrooms, we also had to face the fact that it was too soon to expect measurable student impact from the program. In fact, during the first year of implementation, it is unlikely that many teachers will have reached the stage of being expert, thoughtful users of the new technology. At the end of our study year, our field researchers rated just 16 of the 43 ALS-EL teachers as "high" implementers, and 7 of these were concentrated in just one of the study sites. Again, there is the problem of aggregating outcomes across sites and classrooms, when some teachers are less sophisticated users of technology than others.<sup>10</sup>

### Implications for Future Studies

Becker (1987) proposes an interesting model for field research on the effects of instructional computing. We'd like to quote the rationale for his model at length:

*Traditionally, field experiments have been conducted in order to study the impact of a specific instructional program implemented in similar ways in a small number of closely situated schools. In order to insure faithful implementation of the instructional program, close personal supervision of the sites by the investigator has been the norm. As a result, studies have been conducted in close proximity to the investigator's place of work, and they have tended to involve few schools, teachers, and classes. Often those studies have been of short duration, further limiting one's ability to identify important effects.*

*School district-wide evaluations, the other type of study predominant in the literature reviewed above, are generally longer in duration and involve larger samples of schools, teachers, and classes. But these often lack alternative control-group treatments, and are much less likely to randomly assign classes or schools to treatments, particularly when computers are involved.*

*As a person who has collected and analyzed survey research data for my professional career, I have envied the adequacy of outcome data -- particularly student achievement data -- that field experiments make possible. But I find it limiting to think of measuring effectiveness of computer-based instruction by collecting data from a characteristically unrepresentative and small sample of specially designed programs -- even where alternative treatments and randomized assignment are used.*

*Thousands of teachers and schools pick from the marketplace of computer-related products and ideas in order to accomplish instructional goals. And they produce a wide range of approaches for using computers with their students. To truly understand the effect that computers are having on students in schools, one needs to combine the large domain and representativeness of survey research with the most important characteristics of carefully*

*designed field studies -- appropriate experimental design, over-time measurement of student achievement, and control and measurement of program implementation (p.12).*

He proposes a coordinated series of field experiments involving at least 80 classrooms in sites around the country, focusing initially on just one subject area, mathematics, and one set of grade levels, fifth through eighth. The 80 classrooms would be composed of 40 pairs of experimental and control classrooms, each pair located in the same school and matched in terms of grade level and ability level. The classes in each pair would be randomly assigned to an instructional treatment -- computer use or no computer use. Where possible, students would be randomly assigned to classes as well. All teachers would have to be computer knowledgeable, with at least two years of experience using computers for instruction. There would have to be enough computers available for each computer-using class to have access to one computer for every four students in that class, and the site would have to guarantee a minimum of 30 hours of computer time per student for instruction in mathematics during the year. The specific implementation -- the kinds of software used, the mathematics objectives taught, and the social organization of computer use -- would be decided by the school and the teacher, and therefore would vary substantially from site to site. The only other structural requirement would be that a similar curriculum with similar "opportunities to learn" be provided for students in the control class in the same school.

Becker proposes collecting data systematically to describe the variations in conditions and implementation approaches in each site; and he proposes using both standardized mathematics performance measures across all sites and curriculum-specific measures to match the content of instruction at each site. His design is similar in this respect to the model for technology assessment developed at UCLA for the evaluation of the Apple Classroom of Tomorrow (ACOT) project (Baker, 1987, 1988; and Baker & O'Neil, 1987; as cited in Baker, Herman, & Gearhart, 1989).

The UCLA model emphasizes collaborating with site personnel to develop evaluation goals, designs, and measures, *"providing uniform data collection strategies and measures across the diverse ACOT sites, but reserving places for interests, measures, and effects unique to each site"* (Baker, Herman, & Gearhart, p.4). In this way, they hoped to develop findings that could be generalized beyond the specific context and goals of a single school site, while also providing evaluative data relevant to improving the ACOT program in each site.

The desire for more generalized findings led the UCLA researchers to press sites toward adopting uniform measures of student achievement. Yet, they discovered, their ability to draw conclusions from aggregated, cross-site data was compromised by the diverse contexts and goals of schooling in each site, including wide-ranging grade levels among ACOT classrooms. Given the great diversity in sites, they found it was virtually impossible to control for alternative explanations of findings, i.e., to know whether differences in outcomes were due to differences in teacher quality, overall instructional approach or curriculum, student background, or some other factor. Furthermore, where they succeeded in getting the sites to adopt common measures, they faced the lack of congruence between these measures and local site objectives.

The ACOT program involves much greater diversity in grade levels and subject areas than our study has involved or than Becker expects to include in his field studies program. However, our own study shows that tremendous diversity in goals and approaches can be found even when one is dealing with a fairly narrow grade span, primarily kindergarten or first grade, and a single area of the curriculum -- language arts, and even when teachers have been given the same configuration of hardware, peripherals, and software for computer-based instruction.

Becker seeks to maximize the variation in computer-based approaches under study, though within a single subject area and a small span of grades. The UCLA team -- at least in their writings for the 1989 AERA conference -- seemed to lean toward *reducing* the variation even further than by limiting the grade levels and subjects involved.

*Perhaps it is our university perspective that wishes for a less complex world or at least less complex contexts in which to study ACOT. But questions like the following are hard to avoid. How many resources should be devoted to studying how technology works in such contexts? Should there be an attempt to reduce some of the variation? Should any ACOT resources be devoted to school-level development? Or to demonstrating that a specific ACOT program, under ideal conditions, results in any of the following: in better achievement of existing objectives, acquisition of new and wonderful concepts and skills, and the development of motivated learning? (Baker, Herman, & Gearhart, pp.12-13).*

So, which should it be? Should there be many more, single-site, carefully controlled studies of computer-based instruction, with outcome measures designed to match local objectives? Or, should there be more multiple-site studies that maximize the variety of conditions and circumstances under which the effects of computers can be assessed, combining common measures with content-specific local measures?

We think it could be very worthwhile to pursue a series of *site-specific* studies under one major study umbrella. We see this as an interactive, collaborative effort involving research teams that could be based in several different universities, research centers, or other R&D organizations, and school district teams from around the country. An evaluation team would be assigned to work with each site collaboratively to develop site-specific evaluation strategies. A key feature of the effort would be teacher involvement in defining the goals of computer-based classroom instruction and also choosing or developing indicators of goal achievement. Uniform measures and procedures could be developed for use only in sites that shared the same goals, thus facilitating the meta-analysis of site-specific research findings.

This sort of plan would require much more frequent on-site presence for the evaluation researchers and much closer working relationships with district and school administrators and teachers. School site personnel would have to be willing and able to spend time in a collaborative project, organizational conditions would have to favor collegial communication and shared decision making, and resources would need to be available to support the project over *at least* a three to five year period. We think the investment would be rewarded by research findings that are more immediately useful to the local sites as well as by more general findings that could be developed

through meta-analysis. Because the local sites would benefit directly from the research in their schools, it is possible that some resources for the effort could be obtained from local sources, either public or private.

### Summary

While stressing yet again the complexities of research on computer-based instruction, we want to leave the impression that continuing these evaluation efforts is very important. It is certainly possible to conceive of computers as tools for transforming the teaching and learning process, making it more efficient and effective, and even developing new concepts of the standards and goals that students may reach. However, the money spent on computers could be thrown away if we do not develop a better understanding of how computers can be used to support new ways of teaching and learning.

Our study, though limited in some ways, nevertheless has a variety of contributions to make to the advancement of research methods for technology assessment. Probably the most important contribution we have to make is what we have learned about what influences the successful implementation of computers in classrooms at the early elementary level. Also, our findings about operational variations and levels of implementation quality are important reminders of the need to allow time for major changes to take place and also to know what is implemented before passing judgment on technological innovations. Finally, although our ability to identify student impacts was constrained by the short-term nature of our study design, our classroom observations and interviews with teachers and principals have resulted in hypotheses concerning types of impact that should be examined in more rigorous studies.<sup>11</sup>

## Notes

1. The recommended hardware configuration for the Apple Learning Series: Early Language is one Apple IIGS 512K RGB Professional System; two Apple IIe Color Classroom Systems; and one Apple ImageWriter II Printer. Peripherals to support the ALS-EL program are included in the ALS-EL package; they include two speech synthesizers from Street Electronics, an Echo+ and an Echo IIb; the Touch Window from Personal Touch Corporation; and the Muppet Learning Keys from Sunburst Communications.
2. ALS-EL includes some of the most highly recommended software available from commercial publishers of software for young children. [For an independent rating of early childhood software, see Buckleitner (1989).] Muppets on Stage, Muppetville, and Muppet Word Book (from Sunburst Communications) provide practice in letter, number, shape, and color recognition; matching of upper- and lower-case letters; consonant and vowel identification; and other classification skills. Touch 'N Write (also from Sunburst) uses the touch screen to provide practice in letter formation. Sound Ideas: Consonants, Sound Ideas: Vowels, and Sound Ideas: Word Attack (from Houghton Mifflin) make use of electronic voice synthesizers to help students develop symbol-sound correspondence and word attack skills. ColorMe (from Mindscape) and Where Did My Toothbrush Go? (from D.C. Heath's Explore-A-Story Series) allow children to create and explore combinations of graphics, letters, and words. Talking Text Writer (from Scholastic Software) and the writing slate in the Muppet Word Book enable children to express their ideas in writing, see their own words in print, and share them with others. Talking Text Writer also includes a sound component, which lets children hear the letters, words, or texts they have written.
3. Maddux (1988) describes two general kinds of educational software: tutorial or drill and practice software, which he labels Type I; and word processing, simulation, or programming software, which he labels Type II. With Type I software, he says, "*user involvement is relatively passive, rote skills are emphasized, and control of the interaction between child and computer is largely predetermined by the software developer.*" With Type II software, "*user involvement is relatively active, higher-order thinking skills are emphasized, and control of the interaction between child and computer is largely in the hands of the child.*" Maddux believes both types of software are useful, but that investments in computers can be justified only if teachers emphasize Type II applications. ALS-EL, as noted in Footnote #2 above, includes both types of software.
4. Teachers in the two southern sites received another day of training in 1988-89. This was the extent of Apple Computer's involvement with the sites, since they wanted to demonstrate the program's effectiveness under conditions similar to those that would occur naturally in schools that purchased the package.
5. The CBAM dimensions are described in detail in Gene E. Hall and Shirley M. Hord, Change in Schools: Facilitating the Process. Albany, NY: State University of New York, 1987.

The Stages of Concern (SoC) dimension addresses *"how teachers or others perceive an innovation and how they feel about it."* CBAM predicts that individuals involved in a change effort will progress through seven Stages of Concern. These range from early "self" concerns, which are focused on the implications of the change for the individual's own role or status in the organization and his or her need to be informed, re-think personal orientations, or change personal practices; to "task" concerns, which focus on the logistics of using the innovation and include issues related to efficiency, organizing, managing, scheduling, and juggling demands on time; to "impact" concerns, which focus on the consequences of the innovation for students and ways that impact might be increased.

The Levels of Use (LoU) dimension describes *"what a teacher is doing or is not doing in relation to the innovation."* Three levels of nonuse and five different use levels have been identified in the CBAM research: nonuse, orientation, preparation, mechanical use, routine use, refinement, integration, and renewal. We did not assess this dimension directly, because we felt that doing so accurately would require more training for field researchers than we could provide in this study.

The Innovation Configurations (IC) dimension *"focuses on describing the operational forms an innovation can take."* CBAM generally assumes the existence of a program "archetype" or several acceptable variations against which an individual user's use of an innovation can be compared. ALS-EL, however, has been promoted as a flexible tool that can be adapted to fit differing teaching approaches and language arts curricula. There are so many flexible components to ALS-EL that we not only found it difficult to define a discrete set of acceptable operational variations, but also had trouble discerning any clear patterns in how ALS-EL was used in classrooms. Instead, we found a wide range of variations, as described in another of our AERA presentations. [See Glen Harvey, Diane Kell, and Nancy Gadzik Drexler (1990), "Implementing Technology in the Classroom: Paths to Success and Failure."]

6. Our proposal to Apple Computer suggested a Phase II contract to follow Phase I, and we still had hopes this phase would be funded well into the 1988-89 school year. If we had known that our study would be funded for just one year, we probably would not have devoted such a large share of our study resources to measuring student outcomes, nor would we have asked teachers to devote so much of their time to furthering this goal.
7. For more details on the Sulzby method of eliciting and analyzing children's writing and rereading, see Sulzby, Barnhart, and Hieshima (undated); and Sulzby (1988).
8. Six "focal" students per classroom were randomly sampled from student groups representing high, medium, and low language arts ability as judged by the teacher.
9. Teachers who participated in the project had these roles and responsibilities:

- Administering the TERA to 12 students in the fall and spring. The TERA requires testing each child individually. Test times vary, depending on student age and ability, but usually the TERA can be completed in 15 minutes or less.
- Sending samples of the children's written work, with a label affixed to each sample to identify the student and describe the writing context; and providing a transcript of the student's writing, if necessary to assist in coding the sample.
- Allowing field researchers to observe in their classrooms four times during the year, and allowing them to take focal students out of their classrooms for the Sulzby Writing and Rereading Task on two or three different occasions.
- Being interviewed for 45 minutes to an hour during each field visit.
- Responding to three written survey questionnaires -- in October, March, and May.

Most teachers in most sites were extremely cooperative, for which we are very grateful. We offered up to three days of substitute pay for each teacher to help offset the time required to complete these tasks. Administrators in some sites took advantage of this offer. In other sites, they did not respond or they negotiated for direct payments to teachers or to school personnel who assisted them with study tasks.

10. We developed a rough scale of implementation quality for our study, based on field researchers' assessments of teachers' performance in five areas. We asked field researchers to respond to 11 statements concerning each teacher's use of ALS-EL by assigning a value from 1 to 7, where 1 means the statement was "not at all true" and 7 means the statement was "very true" of the teacher. From those statements we chose 5 that our team of researchers felt were especially important to determining the quality of implementation:

- uses a variety of ALS-EL software components;
- actively plans for curriculum integration;
- is familiar with *all* programs and makes knowledgeable choices;
- uses with *all* students, or varies use for good instructional reasons; and
- makes full or appropriate use of the hardware and peripherals, including sound synthesizers and printer.

We collapsed the seven-point scale to a four-point (0 to 3) scale and then averaged the ratings to come up with our measure of implementation quality. Teachers with average

ratings of 0.2 to 1.4 were considered "low" implementers (7 teachers); those with average ratings of 1.6, 1.8, or 2.0 were considered "medium-low" implementers (10 teachers); those with average ratings of 2.2 or 2.4 were considered "medium-high" implementers (10 teachers); and those with average ratings of 2.6, 2.8 or 3.0 were considered "high" implementers (16 teachers). All teachers in one site were rated as "high" implementers by the end of the 1988-89 school year, but in every other site the teachers were spread across several levels of implementation quality. At the end of the year, there were one or two teachers rated as "low" implementers in four of the six sites. [For further details on this measure and its correlations with other measures, such as the SoC scales, see Harvey, Kell, and Drexler (1990).]

11. See also Harvey, Kell, & Drexler (1990) and Drexler, Harvey, & Kell (1990).

## References

- Baker, E.L. (1988). Sensitive, sensible evaluation of ACOT (Report to Apple Computer). Los Angeles: UCLA Center for the Study of Evaluation.
- Baker, E.L. (1987). Time to write: The U.S. Study of Written Composition. Los Angeles: UCLA Center for the Study of Evaluation.
- Baker, E.L., Herman, J.L., & Gearhart, M. (March 1989). The ACOT report card: Effects on complex performance and attitudes. Presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.
- Baker, E.L., & O'Neil, H.F. (1987). The evolution of the ACOT evaluation model (Report to Apple Computer). Los Angeles: UCLA Center for the Study of Evaluation.
- Becker, H.J. (1987). The impact of computer use on children's learning: What research has shown and what it has not. Baltimore, MD: Center for Research on Elementary and Middle Schools.
- Berman, P., & McLaughlin, M. (1975). Federal programs supporting educational change: Volume IV: The findings in review. Santa Monica, CA: Rand Corporation.
- Berman, P., & McLaughlin, M. (1978). Federal programs supporting educational change: Volume VIII: Implementing and sustaining innovations. Santa Monica, CA: Rand Corporation.
- Buckleitner, W. (1989). 1989 survey of early childhood software. Ypsilanti, MI: High/Scope.
- Clark, R. (1985). Evidence for confounding in computer-based instruction studies: Analyzing the meta-analyses. Educational Communications and Technology Journal, 33:249-262.
- Compaigne, B.M. (1983). The new literacy. Daedalus, 112(1):129-142.
- Cuban, L. (May 1989). Neoprogressive visions and organizational realities. Harvard Educational Review, 59(2):217-222.
- DeFord, D. (1985, Spring). Validating the construct to theoretical orientation in reading instruction. Reading Research Quarterly, 20(3):351-367.
- Drexler, N.G., Harvey, G., & Kell, D. (April 1990). Student and teacher success: The impact of computers in primary grades. Presented at the Annual Meeting of the American Educational Research Association, Boston, MA.
- Edelsky, C., & Harman, S. (October 1988). One more critique of reading tests -- with two differences. English Education, 20(3):157-171.
- Glennan, T.K. (1967). Issues in the choice of development policies. In T. Manschak, T.K. Glennan, & R. Summers (eds.), Strategies for research and development. New York: Springer-Valdez.
- Gold, D.L. (1988). Early testing said to have long-term negative effects. Education Week, 7(27):6.
- Hall, G.E., & Hord, S.M. (1987). Change in schools: Facilitating the process. Albany: State University of New York Press.

Harvey, G., Kell, D., & Drexler, N.G. (April 1990). Implementing technology in the classroom: Paths to success and failure. Presented at the Annual Meeting of the American Educational Research Association, Boston, MA.

Krendl, K.A., & Williams, R.B. (September 1989). The importance of being rigorous: Research on Writing to Read. Unpublished manuscript. Bloomington, IN: Indiana University.

Laboratory of Comparative Human Cognition (February 1989). Kids and computers: A positive vision of the future. Harvard Educational Review, 59(1):73-86.

LaFrenz, D., & Friedman, J.E. (May 1989). Computers don't change education, teachers do! Harvard Educational Review, 59(2):222-225.

Maddux, C.D. (Fall 1988). Research and the state of the art in educational computing. Computers in the Schools, 5(2): .

Meisels, S.J. (April 1989). High-stakes testing in kindergarten. Educational Leadership, 46(7):17-22.

Pearlman, R. (June 1989). Technology's role in restructuring schools. Electronic Learning, 8.

Perelman, L.J. (September 1988). Restructuring the system is the solution. Phi Delta Kappan, 20-24.

Roblyer, M.D., Castine, W.H., & King, F.J. (1988). Assessing the impact of computer-based instruction. Computers in the Schools, 5:3-4.

Shepard, L.A. (April 1989). Why we need better assessment. Educational Leadership, 46(7):4-9.

Starr, P. (1983). The electronic reader. Daedalus, 112(1):143-156.

Sulzby, E. (1988). A study of children's early reading development. In A.D. Pellegrini (ed.), Psychological bases for early education, pp.39-75. John Wiley & Sons.

Sulzby, E., Barnhart, J., & Hieshema, J. (in press). Forms of writing and reading from writing: A preliminary report. In J. Mason (ed.), Reading/writing connections: An instructional priority in elementary school. Boston: Allyn & Bacon.

Sulzby, E., & Teale, W.H. (1985). Writing development in early childhood. Educational Horizons, 64:8-12.

Teale, W.H. (November 1988). Developmentally appropriate assessment of reading and writing in the classroom. Elementary School Journal, 89(2):173-184.

Valencia, S.W., Pearson, P.D., Peters, C.W., & Wixson, K.K. (April 1989). Theory and practice in statewide reading assessment: Closing the gap. Educational Leadership, 46(7):57-63.

Weir, Sylvia (February 1989). The computer in schools: Machine as humanizer. Harvard Educational Review, 59(1):61-73.